

Om bruk av *Oslo-testen* for å undersøke oslomålet

Av Karine Stjernholm og Ingunn Indrebø Ims

The whole is more than the sum of the parts.
Aristoteles, *Metaphysica*

Internett har åpna nye muligheter for å nå fram til store mengder mennesker raskt, og forskere kan få tilgang til svært store datasett på en svært lite ressurskrevende måte. Denne tilgangen til store datamengder åpner for nye metodologiske muligheter og utfordringer. I denne artikkelen sammenlikner vi data fra Oslo-testen, en elektronisk spørreundersøkelse med over 115 000 svar, og Norsk talespråkskorpus – Oslodelen (NoTa), som inneholder opptak med 166 informanter fra Oslo-området. I tillegg drøfter vi metodologiske utfordringer knytta til elektroniske spørreundersøkelser, og vi ser på hva slags teoretisk rammeverk vi kan tolke dataene inn i.

1 Innledning

Våren 2010 ble Oslo-testen (heretter OT) publisert på *Aftenpostens* nettsider. I to måneder var den tilgjengelig på nettet og samla inn om lag 115 000 besvarelser fra respondenter over hele landet. I utgangspunktet hadde testen form som en klassisk spørrelisteundersøkelse, det vil si at respondentene ble presentert for noen utvalgte språklige variabler fra oslomålet, og ble bedt om å rapportere om hvilke variant av variabelen de selv mente de brukte, før de skulle gi en vurdering av de samme variablene på en femtrinnskala. Avslutningsvis ble respondentene bedt om å legge igjen informasjon om alder, bosted, oppvekststed og utdanningsnivå, i tillegg til at de fikk noen spørsmål med åpne svaralternativ. Det som skiller OT fra en tradisjonell spørrelisteundersøkelse, er at den var elektronisk og tilgjengelig via Internett, og at respondenten ble oppsøkt via en publiserings-

kanal (i dette tilfellet *Aftenposten*) – og denne framgangsmåten er fremdeles ikke veldig vanlig i (språk-)vitenskapelig sammenheng.

Internett har åpna en ny mulighet til å nå fram til utrolig mange mennesker på en svært lite ressurskrevende måte. Nye tekniske hjelpemidler har sørget for at langt flere personer umiddelbart er tilgjengelige for undersøkelser som dette, samtidig som disse nye tekniske verktøyene også gjør det mulig å behandle større datasett enn det som har vært mulig tidligere. Denne tilgangen til store datamengder åpner for nye metodologiske muligheter, men også utfordringer. I denne artikkelen vil vi diskutere både mulighetene og utfordringene ved å først plassere OT inn i den klassiske, tradisjonelle spørreliste-tradisjonen den er en del av. Videre vil vi vise hvordan Surowiecki (2004) mener at nye tekniske muligheter åpner for nye muligheter for masseinnsamling av data, og argumenterer for at slike masseinnsamlinger representerer en ny form for diversitet og individualitet i datasettet, og vi vil se på hvilken relevans en slik argumentasjon har for analysene av dataene fra OT. Deretter vil vi gå nærmere inn på skepsisen mot denne typen innsamling og metodikk, hovedsakelig fremmet i Hårstad (2008 og 2010). Endelig er et av hovedaspektene ved denne artikkelen å sammenlikne respondentenes selvrapportering og vurderinger i OT med faktiske talespråksdata. Det sentrale spørsmålet i en slik sammenlikning er i hvor stor grad respondentenes selvrapporterte språkbruk samsvarer med faktisk språkbruk.

2 Spørreliste-tradisjonen og ny teknologi

Spørrelisteundersøkelser har lange tradisjoner i lingvistikken. Den første kjente lingvisten som begynte med denne typen spørreliste, var tyskeren Georg Wenker. I 1876 begynte han å sende ut spørreliste til lærere i Tyskland, og endte opp med å sende til omtrent 50 000 stykker, hvorpå 45 000 returnerte utfylte skjemaer (Chambers & Trudgill 1998 [1980]: 15). Skjemaet inneholdt 40 setninger som skulle transkriberes til lokalt talemål, og på den måten avdekke lokale variasjoner, men tallet på responser og variabler ble svært utfordrende å håndtere, og som Chambers & Trudgill (sst.: 16) skriver: “The wealth of the data, not surprisingly, turned out to be a hindrance rather than an advantage.” Helt fram til slutten av 1900-tallet var nettopp det, å prosessere det store datamaterialet som ble samla inn, det største problemet med denne typen undersøkelser. Denne utfordringa satte en grense for kvantiteten, det ble for mye data, og metodene for å behandle dem var for tungvinne.

Det er fremdeles stor interesse for denne dialektgeografiske tradisjonen som Wenker tilhører. Tekstlaboratoriet ved Universitetet i Oslo har kartlagt mikrosyntaktisk variasjon i Norden gjennom å sende feltarbeidere land og strand rundt med spørreskjemaer. Innsamlinga av disse dataene har vært svært ressurskrevende, men i motsetning til Wenker er resultatene herfra lett tilgjengelige i en database på nettet, der blant annet et elektronisk kart illustrerer den geografiske variasjonen.¹

Også tidligere har Internett blitt brukt til å samle inn data. Blant annet har den amerikanske lingvisten Bert Vaux samarbeida med Tekstlaboratoriet om å spre undersøkelsen “The Cambridge Online Survey of World Englishes” på nettet siden 2007.² Undersøkelsen er hovedsakelig blitt spredd gjennom jungeltelegrafen på nettet, blant annet gjennom sosiale medier, som Facebook o.l. Gunnstein Akselberg og Stian Hårstad gjorde noe liknende da de brukte nettaviser til å publisere spørrelisteundersøkelser om språk i Bergen og Trondheim. Våren 2004 hadde Akselberg undersøkelsen “Hvor pent snakker du?” eller Bergensertesten på nettsidene til *Bergens Tidende* (se Akselberg 2008), og i 2005 publiserte Hårstad Trondheimstesten på nettsidene til *Adresseavisen* (se Hårstad 2008 og 2010). Også i OT har vi brukt nettet til å samle inn dataene.

3 Teoretiske betraktninger

Nettinnsamlinger kan nå potensielt mange flere personer enn det som har vært mulig tidligere. Innsamlingsmetoden og datasettene som er kommet ut av den, er nye i språkvitenskapelig sammenheng. Erfaringsgrunnlaget fra analyser av masseinnsamlinger er sparsomt, og det fins derfor ikke noe etablert teoretisk fundament vi kan støtte oss på her. I dette avsnittet vil vi gjøre noen teoretiske betraktninger rundt hvordan et materiale som dette kan analyseres. Statistisk analyse er selvsagt relevant i denne sammenhengen, og mye av grunnlaget for teoretiseringa vi gjør, vil ha en basis i det som i statistikken kalles de store talls lov. Den går ut på at når et stort antall tilfeldigheter ses under ett, forsvinner det tilfeldige og erstattes med lovmessighet (Aalen 1998: 48). Samtidig vil det ikke bli gjort noen dyptpløyende statistiske analyser her, noe som kommer av at datasettene vi skal sammenlikne, er svært forskjellige. Videre vil vi støtte oss på den allerede nevnte *New Yorker*-journalisten James Surowieckis argumentasjon i boka *The Wisdom of Crowds* (Surowiecki 2004). Han har stor tiltro til større

gruppers kollektive kunnskap, og han begrunner dette særlig ut fra tre hovedperspektiver som vi vil diskutere her, nemlig:

1. Diversitet
2. Individualitet
3. Teknisk utforming

Det første punktet går ut på at diversiteten i store grupper, det at så mange er representert, sørger for at flere (og mange) perspektiver representeres. Det innebærer at tendenser som i mindre datasett virker marginale, kan bli synliggjort. Et lavt antall informanter vil ikke nødvendigvis kunne avdekke mindre frekvente tendenser, fordi disse vil fremstå som så marginale i små datasett. I store datasett vil slike tendenser synliggjøres i mye større grad, og kan dermed tillegges mer rettmessig vekt enn de ville fått med færre informanter. I OT, som er et slikt stort datasett, tydeliggjøres for eksempel variasjon mellom bydelene i Oslo, siden hver bydel er representert med et betydelig antall respondenter. Bydelsvariasjon kommer ikke til å bli drøfta spesielt i denne artikkelen, men dataene fra OT kan bidra til å nyansere bildet av Oslo som delt mellom øst og vest (se forøvrig Hanssen mfl. 1978 og Stjernholm 2013). Surowiecki (2004) mener også at en stor gruppe som helhet er mer troverdig enn det han kaller gruppas sterkeste ledd, noe som er avhengig av neste punkt, som er individualitet.

Surowiecki mener at når hver deltaker i en stor gruppe bidrar med noe av seg selv – uten å vite hva de andre bedriver (respondentene i OT svarte på undersøkelsen uten å vite hva de andre respondentene svarte) – unngår man flokkmentalitet, altså at respondentene begynner å kopiere hverandre. I OT betyr dette at vi har fått innblikk i svært mange respondenters individuelle oppfatninger om språklig variasjon i Oslo, som etterpå lett kan settes i system. Det medfører en mulighet til å få et overblikk over ideologisk fellesgods uten at respondentene sammen har kommet fram til noen konsensus.³

Når det gjelder Surowieckis tredje punkt, som er teknisk utforming, vil vi diskutere dette i lys av våre erfaringer med gjennomføringen av OT. I motsetning til problemene Wenker hadde i sitt forsøk på masseinnsamling, har vi i OT hatt mulighet til – ved hjelp av moderne datateknologi – å holde styr på mengden data vi har fått inn, og vi har hatt verktøy til å trekke ut den relevante informasjonen fra dataene våre. Her har vi tolka Surowiecki

dit hen at massekvantitet kun er nyttig om man har tilgang til de rette (elektroniske) hjelpemidlene for å hente ut essensen i datamaterialet.

Når vi benytter oss av argumentasjonen til en økonomijournalist i en sammenheng som dette, kommer det nettopp av at erfaringsgrunnlaget fra liknende undersøkelser er såpass tynt at vi trenger nye tanker for å kunne tolke dataene på en adekvat måte. Kvantitative data i lingvistisk forskning har tidligere dreid seg om et antall respondenter som ligger fjernt fra hva vi måtte hanskens med i OT. For eksempel skriver Sankoff (1980: 51) at omtrent 150 informanter er tilstrekkelig i kvantitative undersøkelser for å få oversikt over et språksamfunn. Riktignok skriver Sankoff da om kvantitative undersøkelser som innsamling av opptak til korpus, men hun begrunner (det lave) antallet informanter med at ved å ta med flere, blir problemene med databehandlingen for store i forhold til hvor mye ny informasjon man faktisk får tak i. Denne argumentasjonen handler selvsagt ikke om den type undersøkelse eller kvantitet som OT representerer, så dette kan derfor brukes til å illustrere at kvantitet i OT-sammenheng får et helt annet innhold. Sammenlikna med Surowieckis argumentasjon virker det som om den type kvantitet Sankoff skriver om, er svært annerledes enn den typen kvantitet Surowiecki skriver om. Den virkeligheten Surowiecki beskriver, ligger også innafor en helt annen kontekstuell ramme, med tanke på elektroniske hjelpemidler for å nå ut til massene, enn den Sankoff skriver om. Sankoff skriver om en metode der direkte kontakt med informanter er nødvendig, og innafor den forståelsesramma ville ikke en tilnærming til 100 000 respondenter vært gjennomførbar uansett.

I OT ble respondentene bedt om å evaluere variasjon i oslomålet, både individuelt, ved at de skulle velge den varianten de selv mente at de brukte, men også mer generelt, ved at de skulle vurdere de aktuelle variantene på en femtrinns skala.⁴ Preston (1989: 1) hevder at forskere har oversett viktigheten av ikke-lingvisters reaksjoner på språk i sociolingvistikken. Forskning på holdninger til språkbruk har, ifølge ham, med få unntak fokusert på å undersøke folks underbevisste holdninger [covert reactions] til språk, og Preston mener forskningen har mangla fokus på vanlige språkbrukeres persepsjon av språklig variasjon (sst.: 2). Preston mener faktisk at denne typen evalueringer er viktige faktorer for språkendring, de påvirker “[...] the shape of language itself, that is, [they are] important factors in change.” (ibid.). I lys av dette burde folks egne oppfatninger om språk være svært viktige for språkforskerne, men Prestons mening om dette står steilt imot blant annet Bloomfield (1944), som har konstruert ordet “stankos” for

å beskrive ikke-lingvisters meninger om språkbruk. Og som Niedzielski & Preston (2000: vii) skriver: “It is no accident that it looks like a noun form of stank[.]” Motsetningen mellom Preston og Bloomfield illustrerer hvordan skepsisen mot folks egenvurderinger av språkbruk har vært utbredt innen språkvitenskapen (se også Labov 1991[1972]).

4 Hvorfor OT og hva skulle den gjøre?

Den demografiske situasjonen i Oslo er komplisert og i stadig endring, eller som det står i en Fafo-rapport om levekår i Oslo fra 2007: ”Oslo er en film – ikke et foto” (Bråthen mfl. 2007, 36). Byen er, og har til alle tider vært, prega av stor tilflytting, noe som fører til stadige strukturelle endringer (se for eksempel Kjeldstadli 2000 [1994] og Benum 2000 [1994]). Det er også stor grad av intern flytting i Oslo, og flerstedstilhørighet er vanlig i byen. Folk i Oslo bor ikke nødvendigvis bare på ett sted (Aure, Langset & Sørlie 2011: 45). På tross av denne noe flyktige situasjonen viste resultatene fra Talemålsundersøkelsen i Oslo (TAUS) fra begynnelsen av 1970-tallet at språksituasjonen i byen var delt mellom øst og vest (Hanssen mfl. 1978), noe resultatene fra Norsk talespråskorpus – Oslodelen (NoTa) fra begynnelsen av 2000-tallet også har støtta (Johannessen 2008, se også Stjernholm 2013). NoTa består av intervjuer og samtaler med 166 informanter fra Oslo-området, og er tilgjengelig som talespråskorpus på omtrent 900 000 ord.⁵ Det bør også legges til at dette bildet av byen som todelt mellom øst og vest også til stadighet tematiseres i mediene, så todelinga er ofte debattert i den offentlige diskursen.⁶

Hovedsakelig skyldes denne todelinga at industrien i Oslo fra midten av 1800-tallet og fram mot århundreskiftet etablerte seg rundt Akerselva. Mens arbeiderne var bosatt i tilknytning til industrien, var høyere samfunns-lag for det meste bosatt lenger vest (se for eksempel Kjeldstadli 2000 [1994]). Denne geografiske fordelinga var foranledningen til et sosialt skille mellom øst og vest i Oslo, og parallelt med den sosiale utviklinga i byen oppsto et språklig skille. På 1970-tallet var det altså fremdeles mulig å skille mellom to hovedvarianter av oslomål. Kort sagt er østkantmålet en østnorsk dialekt, mens vestkantmålet har sitt opphav fra dansk skriftspråk (se for eksempel Hanssen 2010: 148 og Larsen 1907). Flere undersøkelser har vist at de tradisjonelle opposisjonene mellom øst- og vestkantmålet fortsatt fins, men at de er svekka i omfang. Amund B. Larsen skriver på begynnelsen av 1900-tallet at de to utviklingslinjene i oslomålet konvergerer (Larsen 1907:

13), og seinere er dette også påpekt av blant andre Alnæs (1963), Jahnsen (2001), Opsahl & Røyneland (2009), se også Stjernholm (2013) for en videre drøfting.

OT ble i utgangspunktet til fordi vi var interesserte i å finne ut noe om den helhetlige samtidige språksituasjonen i byen. Samtidig ble det på mange måter et metodeeksperiment, og tanken bak testen ble derfor todelt. Vi ønska både å finne ut noe om den helhetlige talespråkssituasjonen, samtidig som vi ville finne ut noe om hvordan nettet kunne brukes til å samle inn data om språk. For selv om vi sitter på store mengder talemålsopptak fra Oslo, er det usikkert om de alene kan fortelle oss noe om den helhetlige språksituasjonen i byen, jamfør diskusjonen rundt Surowiecki og Sankoff i forrige avsnitt. Ved å utelukkende gjøre mer kvalitative nedslag i byen, kan det være vanskelig å gripe det helhetlige bildet. Ved å nå ut til store mengder osloborgere gjennom OT, hadde vi et håp om at vi skulle lykkes i å få dette helhetlige inntrykket av språksituasjonen i byen, der også understrømmene kunne være synlige, jf. Surowieckis argumentasjon om individualitet.

I tillegg til den tidligere nevnte skepsisen mot ikke-lingvisters oppfatninger av egen språkbruk, er det også en utbredt skepsis blant lingvister mot å bruke nettet som innsamlingskanal for språkdata. Ikke minst har Hårstad, som selv har erfaring med dette, uttrykt at han stiller seg svært kritisk til metoden (se Hårstad 2008 og 2010), noe vi vil drøfte videre i avsnitt 6. Likevel mente vi det var gode grunner til å gå i gang med denne innsamlingen, særlig siden dataene fra NoTa var såpass ferske. Gjennom NoTa har vi tilgang til omtrent samtidige opptak av oslomål. På den måten kan vi sammenlikne svarene fra OT med faktisk språkbruk, og dette er et unikt utgangspunkt for å undersøke forholdet mellom rapportert og faktisk talespråk i Oslo.

5 Oslo-testen – form og innhold

OT samla altså inn til sammen omtrent 115 000 svar fra hele landet – cirka 50 000 av respondentene rapporterte at de bodde i Oslo. Testen var selv-rekrutterende, og ble lagt ut på nettet klokka 23.30 natt til søndag 10. april 2010, og innen klokka 6.30 var den besvart 2500 ganger. Den var også den nest mest leste saken på nettutgaven til *Aftenposten* i hele 2010. Siden Bergenstesten og Trondheimstesten også fikk mange svar, henholdsvis 56 000 og 28 000, var vi forberedt på stor respons, men over 100 000 på noen få uker var mer enn vi hadde regna med.

5.1 Form

Surowiecki (2004) mente at tekniske løsninger er avgjørende for å muliggjøre datapotensialet til store grupper, noe som i høyeste grad gjelder for OT. Vi var opptatt av at testen skulle utformes på en måte som gjorde det lett for respondentene å forstå testen, i tillegg til at vi hadde noen tanker om hvordan respondentene logisk skulle finne sin sti gjennom det språklige landskapet. Hele veien hadde vi i tankene at formen på testen var avgjørende for hvor mange som ville gjennomføre den. Vi samarbeida tett med *Aftenposten* for å lage et testformat som skulle fungere,⁷ og i dette avsnittet vil vi se nærmere på utforminga av OT.

Hele OT var på 26 sider. På de første 22 sidene ble respondentene presentert for 22 språklige variabler fra oslomålet. Variablene var både fonologiske, morfologiske, syntaktiske og leksikalske. Undersøkelsen var anonym, men vi ba respondentene gi opplysninger om kjønn, aldersgruppe, utdanningsnivå, oppvekstfylke (obligatorisk), postnummer, ev. bosted i dag (ikke obligatorisk). Respondentene ble også bedt om å svare på hva de kalte måten de selv snakker på, om de brukte andre språk enn norsk til daglig, og eventuelt hvilke(t) språk dette var.

Figur 1 under viser én (av 22) side(r) med en variabel fra OT. På disse sidene skulle respondentene rapportere hva de mente at de selv sa ved å trykke på den boksen som inneholdt “deres” alternativ, og de skulle også vurdere variantene på en femtrinnskala fra sært–greit under:

Sier du ÅSSEN eller VORDAN?

ÅSSEN har du det? VORDAN har du det? Jeg veksler mellom dem

Klikk for å markere hvor sære eller greie du synes disse ordene er

Åssen Sært Greit

Vordan

Figur 1: Første side med spørsmål i OT: “Sier du ÅSSEN eller VORDAN?”

Som det går fram av bildet, var variantene satt inn i en eksempelsetning, og respondentene skulle velge om de brukte den ene eller andre varianten, eller om de veksla mellom dem. Når respondentene hadde foretatt et valg, kom sært–greit-skalaene til syne. Uavhengig av hva respondentene valgte som “sin” variant, skulle de vurdere begge variantene på sært–greit-skalaene etterpå. Når respondentene hadde fylt ut alle punktene på siden, kom det til syne en knapp der de kunne trykke seg videre til neste side med neste variabel. Det var ingen mulighet for å ikke svare på spørsmålene, alle måtte besvares for å komme seg videre i testen, og det var heller ikke mulig å gå tilbake til sider man hadde vært gjennom. Det var kun én retning i testen.

5.2 Innhold

Mordal (1989: 53) mener spørrelisteundersøkelser bør bygge på både teoretisk kunnskap og tidligere erfaring. Hun mener det er viktig og riktig å basere hypoteseutvikling til spørreundersøkelser på egne erfaringer, og hun vektlegger viktigheten av å fange opp hva som er det vesentlige i den aktuelle problemstillinga. Den vesentlige problemstillinga i denne sammenhengen er Oslos velkjente todeling mellom øst og vest.⁸ Når du spør oslofolk hva som skiller de to ulike varietetene i byen, dukker ofte begrepet a-ender opp. Dette er et (til tider noe vagt) begrep som vi kan se eksempler på i utdragene fra NoTa under, der én informant fra vest (181 fra Frogner) og én fra øst (192 fra Stovner) karakteriserer sitt eget språk (våre uthevinger):

Intervjuer: em (...) hvis du skulle karakterisere språket der som du vokste opp (.) hva ville du si om det da?

NoTa 181: fint (.) eller i hvert fall finere enn østkantdialekt det ville jeg vel sagt men e (...) det er jo noe man vokser fra seg etter hvert som man man snakker kanskje penere når man er yngre enn det man gjør når man er eldre

Intervjuer: men hva hva liksom mener du med fint kan du gi noen eksempler?

NoTa 181: ikke a-ender (.) veldig sjeldent a-ender det er ikke “mora mi” men “moren min”

NoTa 192: jeg kommer jo ifra østkanten da så da vil jeg karakterisere det som veldig sånn typisk østkant det vil si (.) a-ender og (.) trykk på forskjellige ord sånn som (...) istedenfor “moren min” så blir det “mora di” ikke sant

“A-ender” er altså en variabel som kan regnes som vesentlig i denne aktuelle problemstillinga, men vi ønska også å finne ut om folks oppfatning av øst- og vestkantmål hadde en tydelig kopling til andre språklige variabler enn de som kan gå inn i sekkebetegnelsen “a-ender”. Diskusjonen og analysene i denne artikkelen er gjort på bakgrunn av de første 11 sidene i testen, som inneholdt 11 variabler. Disse variablene representerer tradisjonell oslovariasjon, i motsetning til de neste 11, som undersøker folks oppfatninger om slang og språklige trekk knytta til oslomål snakka i flerspråklige og flerkulturelle miljø i Oslo. Disse sistnevnte variablene er sentrale i Ims (u.a.). Tabellen under viser de første 11 variantene/variablene vi testa i OT:

Øst	Vest	Kommentar
åssen	vordan	Leksikalsk variasjon
sjøl	sell	Refleksiv
mora di/mor di	din mor/moren din	Syntaktisk/morfologisk
finni	funnet	Perf. part. sterke verb
driti	dritet/dritt	Perf. part. sterke verb
håppa	håppet	Pret. svake verb, 1. kl.
døra	døren	Best. entall, substantiv
fingra	fingrene	Best. flertall, substantiv, M
veit	vet	Diftong/monoftong
gammal	gammel	Bevart/svekket vokal i rotv.
trur	tror	Rotvokal u/o

Tabell 1: De første 11 variablene i OT. (Vi har med vilje unngått å ta stilling til om a-ender i substantiv bestemt form entall bare forekommer ved feminine substantiv. Lødrup (2011) argumenterer for at vestkantdialekten i Oslo ikke har feminint genus, at bøyningssuffikset *-a* ikke er relatert til femininum.)

5.2.1 Skrivemåten

Siden vi skulle bruke et skriftlig medium for testen, hadde vi noen åpenbare utfordringer knytta til at det var talespråksvalg vi ønska informasjon om. Lydfiler ble vurdert som supplement til testen, men vi var redde for at ved å gjøre testen mer teknisk avansert, kunne problemet med frafall bli større, og vi hadde ingen forutsetninger på forhånd for å vite at responsen på undersøkelsen skulle bli så overveldende. Vi hadde for eksempel tanker om at vi kunne nå ut til innbyggere uten internetttilgang hjemme via PC-er på offentlige steder som biblioteker o.l., og vi tenkte det var sannsynlig at disse stedene ikke hadde PC-er med lyd.⁹

Det var fristende å inkludere flere fonologiske variabler i OT som er saliente i språklandskapet i Oslo. Realiseringa av konsonantsambandet *-s/* (se Kristoffersen & Simonsen 2008, Opsahl & Røyneland 2009), realiseringa av *l* (Svendsen 2012, Jahr 1986), eller trykkforholdet i fremmedord med gresk eller latinsk opphav er noen eksempler, men siden variablene skulle presenteres skriftlig, ble disse utelatt fordi vi anså at sannsynligheten for misforståelser var for stor. Flere av variablene er skrevet med det som kan kalles en talemålsnær transkripsjon. Bakgrunnen for å bruke en slik skrivemåte var et forsøk på å utlikne forskjellen mellom varianter som er vanlige i skrift, og de som ikke er det, noe som blant annet er fundert i dette utsagnet fra Johnson (2008: 37):

The context in which a judgement occurs influences it greatly. So if the test sentence appears in a list that has lots of "informal" sentences of the sort that language mavens would cringe at, it may get a higher rating than if it appeared in a list of "correct" sentences.

For eksempel viser et søk i Leksikografisk bokmålskorpus (LBK),¹⁰ en samling med bokmålstekster fra 1985 til i dag, at *hvordan* er mye vanligere i skrift (bokmål) enn *åssen*. Mens *hvordan* er representert 52 935 ganger i korpuset, er *åssen* bare brukt 585 ganger.¹¹ Siden det er et skriftbilde respondentene måtte forholde seg til i OT, var det stor sannsynlighet for at det i altfor stor grad påvirket hvilken variant de valgte. Det er heller ikke uvesentlig at OT ble formidla gjennom *Aftenposten*, som altså er en språklig konservativ avis. Ved å velge en talemålsnær transkripsjon, altså å skrive *åssen* på vanlig måte, men *vordan* uten *h*, mente vi begge variantene kunne oppfattes som markerte, noe vi håpa ville føre til at de stilte mer på lik linje i testen. Tanken var også at dette valget ville føre til den tilleggseffekten at

respondentene hele tiden fikk en påminnelse om at denne testen ikke handla om skriftspråk, og ved å ignorere vanlige skriveregler mente vi å understreke dette. Når vi kommer til resultatene, vil vi vise at det kan se ut som vi har lyktes langt på vei med dette. For eksempel har varianten *sell* blitt evaluert som forventa med tanke på talespråk (*sell* er totalt mer frekvent i NoTa enn *sjøl*, se Figur 6 i avsnitt 7.3). Hvis respondentene utelukkende hadde evaluert skriftbildene, skulle man forvente at en framstilling som for eksempel *sell*, som kan ses som svært avvikende fra skriftbildet, skulle evalueres lavt, og det er ikke tilfellet. Man kunne også tenke seg at skrivemåten potensielt kunne medført misforståelser. Det kunne tenkes at respondentene ikke ville forstå at *sell* er det samme som det vi skriver som *selv*, men dette problemet unngikk vi fordi alle variantene av samme variabel var synlige samtidig. Når respondentene så *sell*, så de samtidig varianten *sjøl*.

I det hele tatt ser det ut til at de spørsmålene vi stiller i OT, besvares som vi ønska på forhånd, det vil si at svarene fra respondentene ser ut til å reflektere talespråklig variasjon, noe vi vil vise i avsnitt 7. Men før vi kommer dit, vil vi se litt nærmere på kritikken som er retta mot denne typen undersøkelser.

6 Kritiske punkter

Som nevnt tidligere er det en utbredt skepsis mot bruk av nettet for innsamling av språkdata. Kritikken mot spørreundersøkelser på nettet gjelder særlig utfordringer knytta til validitet og representativitet. Dette har Hårstad (2008: 138-ff.) drøfta spesielt, og diskusjonen her vil ta utgangspunkt i kritikken hans. Hårstad viser bl.a. til statsviterne Best & Krueger (2004), som er opp-tatt av at brukernes tilgang til Internett kan være varierende. Hårstad tar også for seg problemer knytta til frafall av respondenter, samt utfordringer med at selve testsituasjonen er uoversiktlig for forskeren, og er vanskelig å kontrollere.

I det følgende vil vi gjennomgå disse argumentene. Først vil vi ta en titt på hvem som er på nett i Norge i dag.

6.1 Hvem er på nett?

I Norge hadde 93 prosent av husholdningene Internett i 2010, og 84 prosent hadde bredbåndstilgang.¹² Fra 2001 til 2010 har tallet på avislesere på nett økt fra 10 prosent til 43 prosent.¹³ Visstnok leser flere menn enn kvinner

nettaviser, men fordelinga mellom kjønn i OT er nesten lik, med 50,4 prosent kvinner og 49,6 prosent menn.¹⁴ Når det gjelder aldersfordelinga i OT-materialet, er i gjennomsnitt 12 prosent av befolkninga mellom 13 og 59 år representert i materialet.

Andelen respondenter av Oslos befolkning som har besvart testen, sortert etter aldersgrupper, er fordelt på følgende måte: 12,3 prosent av Oslos befolkning under 20 år har besvart testen,¹⁵ 15,5 prosent av befolkninga mellom 20–29 år, 13,5 prosent mellom 30–39 år, 9,7 prosent av de mellom 40–49 år, 7,2 prosent av de mellom 50–59 år,¹⁶ og 3,2 prosent av befolkninga over 60 år. Tallene er beregna i forhold til folkemengden i Oslo per 1. januar 2012,¹⁷ og her har vi ikke tatt høyde for at noen kan ha besvart testen flere ganger, noe som vil bli diskutert i avsnitt 6.3. De aller fleste respondentene fra Oslo i OT er altså mellom 20–39 år gamle, noe som stemmer overens med tall fra Oslostatistikken (2013), som viser at aldersfordelinga i Oslo avviker betydelig fra landsgjennomsnittet: 37 prosent av Oslos befolkning tilhører aldersgruppa 20–39, noe som er 10 prosentpoeng høyere enn andelen for resten av landet. For aldersgruppa 6–19 år og personer over 80 år er andelen i Oslo derimot under landsgjennomsnittet (se Ims 2013: 49). Siden akselerasjonen i den elektroniske utviklinga er rask, og bruken av datamaskiner og Internett i befolkninga er såpass høy, kan det tenkes at det Best & Krueger (2004) hevder, allerede er noe passé.

6.2 Frafallsproblematikk og kognitiv lønnsomhet

I språkvitenskapelig sammenheng har OT samla inn svar fra en veldig stor gruppe, men det er likevel en viss fare for at respondentene kan representere et noe skeivt bilde av osloborgerne. Eller som Johnson (2008: 35) skriver: “[N]onrandom samples are likely to have some bias in them.” Dette er en åpenbar svakhet med metoden, samtidig hadde vi på forhånd et håp om at dette problemet til en viss grad ville bli kompensert for av et høyt respons-tall. Vi var derfor opptatt av hvordan vi kunne sikre at flest mulig gjennomførte testen, og diskuterte nøye hvordan vi skulle få flest mulig til å svare på hele testen. Testen var som nevnt på 26 sider, så frykten for at respondentene skulle bli lei og avbryte underveis i undersøkelsen, var høyst reell.

Hårstad (2008: 132 og 2010) skriver at det er viktig at en undersøkelse som dette er kognitivt lønnsom, og for at ikke respondentene skal trekke seg ut av testsituasjonen, må den gi mer tilbake til respondenten i form av læring eller underholdning enn det de selv må legge ned av innsats. I OT

fikk respondentene en slags belønning for gjennomført spørreskjema ved at de fikk en språklig diagnose eller merkelapp på slutten av testen, basert på hvilke språklige valg de tok underveis. Belønninga besto i å havne i én av fire kategorier: Tradisjonell Øst, Tradisjonell Vest, Urban Oslo eller Gatesmart. I én av de to første kategoriene kom respondentene dersom de hadde overvekt av former som tradisjonelt er assosiert med tradisjonelt østkant- eller vestkantmål. Urban Oslo havna de i dersom de hadde en blanding av former fra øst og vest, og Gatesmart havna de i om de mente at de for eksempel brukte avvikende kjønn på substantiv eller syntes det var greit med manglende inversjon i undersetninger. Denne kategoriseringa skapte – ikke overraskende – mye debatt. Hårstad (sst.) påpeker at det kan være problematisk å viske ut grensene mellom seriøs forskning og underholdning på denne måten. Men i seg selv er det ikke uvanlig at det gis en form for belønning for å delta i et forskningsprosjekt. Det uvanlige i denne sammenhengen er at innsatsen i selve undersøkelsen påvirka utfallet av belønninga direkte. Samtidig var det åpenbart at all viraken rundt disse kategoriseringene var noe av årsaken til at OT ble så populær. På tross av at det tok relativt lang tid å komme seg gjennom hele testen for å nå resultatet, ble altså testen gjennomført 115 000 ganger, og det at man for eksempel kunne dele resultatet sitt på Facebook og diskutere det med andre, var trolig avgjørende for rekrutteringa til testen.

En rent statistisk generalisering kan bare foretas til en populasjon der utvalget man tar, oppfattes som et tilfeldig utvalg av akkurat denne populasjonen (Aalen 1998: 199). Siden respondentene i OT ikke er et tilfeldig utvalg av en gitt populasjon, kan vi selvsagt ikke gjøre en rent statistisk generalisering over disse dataene. Likevel er det ikke utvalget i seg selv som er den største trusselen mot å generalisere over disse dataene, siden det store antallet besvarelser gjør at de store talls lov gir oss relativt godt grunnlag for å generalisere. Det største problemet i denne sammenhengen handler mer om hva vi generaliserer over.¹⁸ Det store antallet besvarelser i OT bidrar til å gjøre andre utfordringer knytta til testformatet mindre, som utfordringer ved testsituasjonen, som er det neste vi vil diskutere.

6.3 Testsituasjonen

Hårstad (2008: 140–141) har vektlagt at utfordringer tilknytta selve testsituasjonen er et sentralt problem ved internettundersøkelser. For det første vet forskeren lite om hva som skjer i testsituasjonen, om for eksempel hvilke tolkninger respondentene har lagt til grunn for svare de har gitt, for

det andre har forskeren lite kontroll over eventuelle ønsker om å manipulere testresultatene. For å ta det siste først – dette gjelder all innsamling av språkdata, og trolig kan betydninga av dette problemet sies å være omvendt proporsjonalt med antall besvarelser. Desto flere besvarelser man samler inn, desto vanskeligere er det å manipulere testresultatene, siden effekten av variasjon reduseres ved gjentakelser (Aalen 1998: 199). Det er dessuten mulig å gjøre dette problemet mindre ved å luke ut IP-adresser som er registrert flere ganger, noe som kan tyde på at én person har besvart undersøkelsen flere ganger. Etter å ha silt ut ufullstendige svar fra OT, satt vi igjen med 100 015 svar. Ved å bare ta med førstegangsregistrering av IP-adressene, satt vi igjen med 90 782 stykker. De IP-adressene som er registrert flere ganger, kan naturligvis være et resultat av useriøse besvarelser, men det kan også tenkes at enkelte adresser for eksempel representerer en hel skoleklasse som har tatt testen på samme datamaskin på skolen, eller at flere mennesker som ikke har tilgang til datamaskin eller Internett hjemme, har tatt testen via samme datamaskin på et offentlig bibliotek eller liknende. Det kan også tenkes at flere av de eldre respondentene (som er de som kanskje er vanskeligst å nå i slike undersøkelser) ikke besitter én datamaskin per person i husstanden, men heller deler samme maskin. Det kan derfor være at de IP-adressene som er registrert flere ganger, faktisk sørger for en bedre representativitet i datasettet. Alle eksemplene fra OT som analyseres her, har også vært analysert på bakgrunn av det datasettet der IP-adressene kun er registrert én gang, og det er ingen forskjell av betydning mellom dem, noe vi vil komme tilbake til seinere i artikkelen. Det har altså ingenting å si for analysen om det ene eller det andre datasettet legges til grunn, men vi vil likevel hevde ut fra argumentasjonen over at det er et bedre metodologisk valg å inkludere alle besvarelsene framfor å trekke ut noen.

Det andre problemet er at vi vet lite om hva respondentene legger i vurderingene sine. Vi vet for eksempel lite om i hvilken grad vi har testa skrift- eller talespråk, et skille som er betydningsfullt for en språkforsker, men kanskje mer underordna for mange av respondentene. Vi vet heller ikke om respondentene svarer det de tror de sier, det de skulle ønske de sa, eller det de tror andre mener de bør si. Det kan dessuten tenkes at respondentene har blitt unødig forvirra av at de først skulle rapportere egen språkbruk, og deretter ble bedt om å vurdere alle variantene, noe som impliserer at de eventuelt må vurdere varianter som de selv ikke bruker, på en skala. Det er stor forskjell på å vurdere egen språkbruk og det som kan tolkes som andres språkbruk.

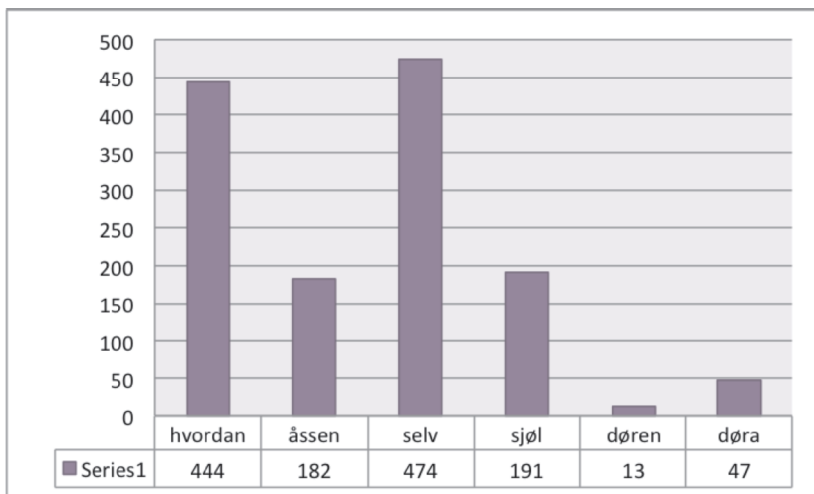
Mordal (1989: 39) påpeker at det ikke bare er vanskelig å avsløre hva folk mener, men også hvor sterk deres mening er, og om de i det hele tatt har noen mening om det vi spør om. Også denne problemstillinga er aktuell i all type innsamling av (språk)data, men i OT blir dette mindre viktig med tanke på hva som er hovedmålet for undersøkelsen, nemlig å kartlegge hovedstrømninger. Eventuell variasjon i fortolkninga av spørsmåla veies opp av antallet responser, siden robuste tall tåler ekstremverdier godt (Aalen 1998: 195).

7 OT og faktisk språkbruk

Som nevnt tidligere mener Preston (1989) at språkforskere har oversett viktigheten av folks bevisste holdninger til språk, og han mener også at disse bevisste holdningene er viktige i språkendringsprosesser. Dette står på et vis i opposisjon til hva undersøkelser i for eksempel Danmark har vist, der språkutviklinga ikke ser ut til å følge de bevisste [overt] holdningene til språk, men heller de underbevisste [covert] holdningene (se Kristiansen & Coupland 2011). Ut fra utviklinga i Danmark skulle vi forvente at resultatene i OT i større grad reflekterte et prestisjehierarki, altså at respondentene ville være mer opptatt av å velge varianter i OT som de mente representerte det de oppfatta som det beste språket framfor å egentlig reflektere over hvilke variabler de selv bruker. Dette kan også ha innvirkning på språkendringsprosesser i den forstand at man tenker at folk vil framelske varianter de opplever som prestisjefulle i en eller annen form – og det kan de jo i og for seg også ha gjort. Men på bakgrunn av dette var det knytta stor spenning til om resultatene fra OT ville samsvare med talemålsopptakene i NoTa.

I dette avsnittet vil vi presentere en slik sammenlikning, representert ved de to variablene (*h*)*vordan/åssen* og *sel(l)v/sjøl*. Det kunne i høyeste grad vært interessant å sammenlikne flere av variablene fra NoTa, men dette viste seg å ikke være så enkelt som vi i utgangspunktet hadde tenkt. Bare de variablene i OT som kan kalles grammatiske, kan sammenliknes direkte med korpuset. For eksempel er variablene som er ment å undersøke a-ender bare én leksikalsk forekomst,¹⁹ mens variablene *vordan/åssen*²⁰ og *sell/sjøl* bare representerer seg selv, eller sin egen grammatiske funksjon. I Figur 2 under har vi vist at sammenlikningsgrunnlaget fra NoTa er svært forskjellig for *døra/døren*, som altså er en leksikalsk forekomst, og for *hvordan/åssen* og *selv/sjøl*. Både *hvordan/åssen* og *selv/sjøl* er relativt fre-

kvente i NoTa, mens et leksikalsk ord som *døren/døra* har færre treff (se for øvrig Hanssen 1986 for en drøfting av ordforråd i et talemålsmateriale):



Figur 2: Illustrasjon av frekvensen til *hvordan/åssen*, *selv/sjøl* og *døra/døren*.

De fire første søylene viser altså antall treff på to og to varianter som i par representerer en grammatisk kategori, mens de siste to søylene kun representerer en leksikalsk enhet. Ut fra det vi ser her, er det ikke særlig sannsynlig at utbredelsen av *døren/døra* i NoTa kan gi et fullgodt bilde av forholdet mellom *-a/-en* i bestemt form entall av substantiv i oslomålet, noe som varierer fra ord til ord (Stjernholm 2013), og det samme gjelder for flesteparten av variablene vi har testa i OT. Det er fullt mulig at dataene fra OT kan bidra med nyttig informasjon om disse variablene, men i forholdet mellom NoTa og OT og variablene *(h)vordan/åssen* og *sel(l)v/sjøl*, er det en mer direkte parallell.²¹ Før vi begynte analysearbeidet, måtte vi avgrense hvem respondentene i OT var, det vil si hvem vi skulle bruke som grunnlag for analysene, og hvem av informantene i NoTa vi skulle sammenlikne dem med.

7.1 Valg av respondenter/informanter

Vi har valgt å kalle testpersonene fra OT respondenter og testpersonene i

NoTa for informanter. Sånn vi ser det, har informantbegrepet mer kvalitative konnotasjoner enn det respondentbegrepet har. I NoTa er informantene fra Oslo og omegn, altså Oslo-regionen (se diskusjon i Stjernholm 2013). NoTa inneholder informanter som kunne være født og oppvokst i pendleravstand til Oslo, men det ble ikke aktivt rekruttert informanter utenbys fra.²² Resultatene fra OT som presenteres her, inneholder kun svar fra de 51 354 respondentene som oppgir at de bor enten øst eller vest i Oslo by.²³ Det vil si at vi har unnlatt å ta med 2371 respondenter som kun har oppgitt “Oslo”, og vi har også holdt 39 respondenter som oppgir at de bor i Marka i Oslo, utenfor, fordi Marka i Oslo omkranser hele byen og ikke nødvendigvis representerer ett sted. Dessuten er det så få respondenter som har oppgitt å bo der, at det å ta dem ut ikke vil påvirke tallene i noen betydelig grad. Disse 2410 respondentene representerer 4,5 prosent av det totale antall respondenter fra Oslo (53 764), og er utelatt for at vi skal kunne sammenlikne dataene med NoTa-korpuset, der informantene er delt etter bosted i vest eller rest.

Regionsvariabelen i NoTa er altså delt i vest og rest. De vestre bydelene var Frogner, Nordre Aker, Nordstrand, St. Hanshaugen, Ullern og Vestre Aker, bydelene i rest var Alna, Bjerke, Gamle Oslo, Grorud, Grünerløkka, Sagene, Stovner, Søndre Nordstrand og Østensjø. Akershus-kommunene Asker og Bærum ble regna som vest, mens de andre Akershus-kommunene ble regna som rest²⁴ (se sst. for en drøfting). Siden bakgrunnen til NoTa-informantene også er svært sammensatt, ville det blitt noe søkt å forsøke å plukke OT-respondenter med identisk bakgrunn, så vi har satt en grense for respondenter fra OT som oppgir at de bor i byen. Respondentene i OT har også rapportert hvilket fylke de er vokst opp i, og blant OT-respondentene som oppgir å bo i Oslo, er alle fylker i Norge representert som oppvekstfylker. 61 prosent av OT-respondentene som bor i Oslo, oppgir Oslo som oppvekstfylke. Oslo har til alle tider vært en innflytterby (Kjeldstadli 2000 [1994]), så at mange oppgir å være utenbys fra, var ikke overraskende (se Stjernholm 2013 for diskusjon av flyttestrømmer, stedsideologi og språkvalg i Oslo).

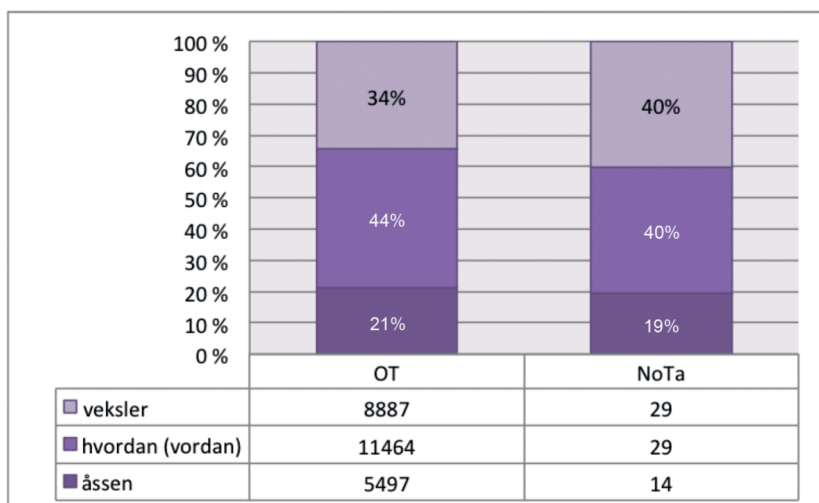
Resultatene fra NoTa var basert på et søk der alle korpusets 166 informanter er med, men vi valgte å ta ut informant nummer 039, som er fra Vestfold og helt klart representerer et annet dialektområde enn Oslo. Derfor er det 165 informanter til sammen i disse analysene.²⁵ NoTa-korpuset har et utvalg på 144 av de 166 informantene som er ment å være et representativt utvalg osloborgere. Dette utvalget er utelukkende teoretisk re-

presentativt, på den måten at de 144 informantene er likt fordelt over variablene kjønn, alder, bosted og utdanning. Informantutvalget er ikke statistisk representativt på den måten at det er gjort et utvalg av informanter basert på en avgrensa populasjon, og at informantene representerer denne populasjonen (se for eksempel Aalen 1998). Det er for eksempel ikke tatt høyde for at utdanningsnivået varierer med hensyn til bosted, og heller ikke at andelen innvandrere varierer geografisk. Fordi vi mener at det ikke fins et godt fundert grunnlag å plukke ut disse 144 informantene på, har vi altså inkludert alle NoTa-informantene i disse resultatene med unntak av 039. NoTa-informantene er kategorisert etter stedstilhørighet i henholdsvis vest og rest, det vil si at informantene er jevnt fordelt etter om de har bodd lengst i vest eller rest. Men som nevnt er for eksempel flerstedstilhørighet svært vanlig i Oslo (Aure, Langset & Sørli 2011: 45), og siden vi mener at valg av bosted er viktigere for talemålssituasjonene i byen enn hvor informantene har bodd lengst (se diskusjon i Stjernholm 2013), har vi sortert søkene etter bosted, altså hvor informantene bodde på opptakstidspunktet. 80 informanter representerer rest, og 85 informanter representerer vest. Som vi skal se, var resultatene fra NoTa oppsiktsvekkende like selvrapporteringene fra OT.

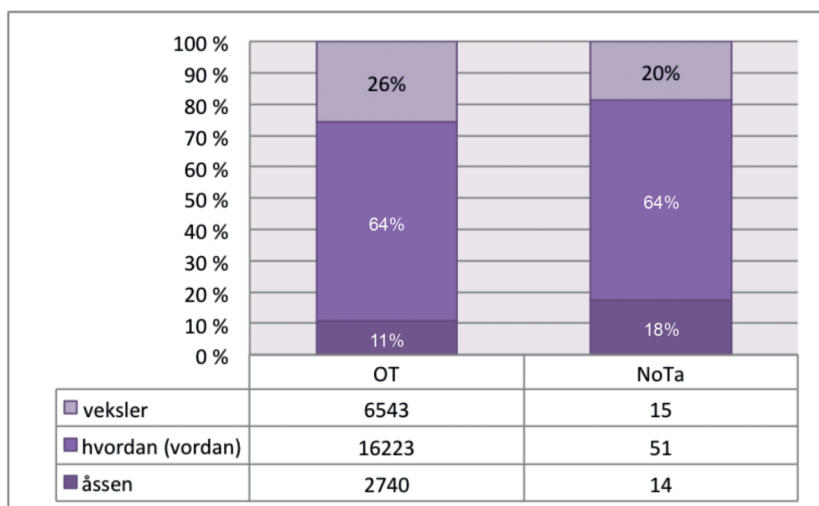
7.2 Første par ut: *åssen* og *(h)vordan*

Den første variabelen vi skal se på, er *åssen* og *(h)vordan*, og vi begynner å se på regionen rest. I figur 3 under ser vi tallene fra de respondentene i OT som har oppgitt å bo i østlige bydeler, i tillegg til sentrum, sammenlikna med informantene i NoTa som bor i rest.

Fordelinga av de to variantene er veldig lik. I OT rapporterer 21 prosent at de sier *åssen*, mens 19 prosent av informantene i NoTa bare bruker *åssen*. 44 prosent av respondentene i OT mener de bare bruker *vordan*, mens 40 prosent av informantene i NoTa bare bruker *hvordan*. At så mange som 34 prosent av respondentene i OT var klar over at de veksle mellom *åssen* og *hvordan*, var overraskende. Og når vi ser hvor mange som faktisk veksler i NoTa (40 prosent), ligger dette overraskende nært hvor mange i OT som faktisk mener at de veksler. Trolig veksler de fleste osloborgere noe mellom *åssen* og *hvordan* selv om det ikke nødvendigvis er dokumentert i NoTa at alle gjør det, men det at tallene fra OT og NoTa ligger så nært hverandre, tyder på at respondentene her gir et realistisk bilde av distribusjonen til *åssen/hvordan* i språklandskapet Oslo. Den like fordelinga mellom NoTa og OT er også synlig i vest, selv om den ser litt annerledes ut der, som Figur 4 viser:



Figur 3: Fordelinga av (h)vordan og åssen for respondentene fra øst i OT²⁶ og informantene rest i NoTa.

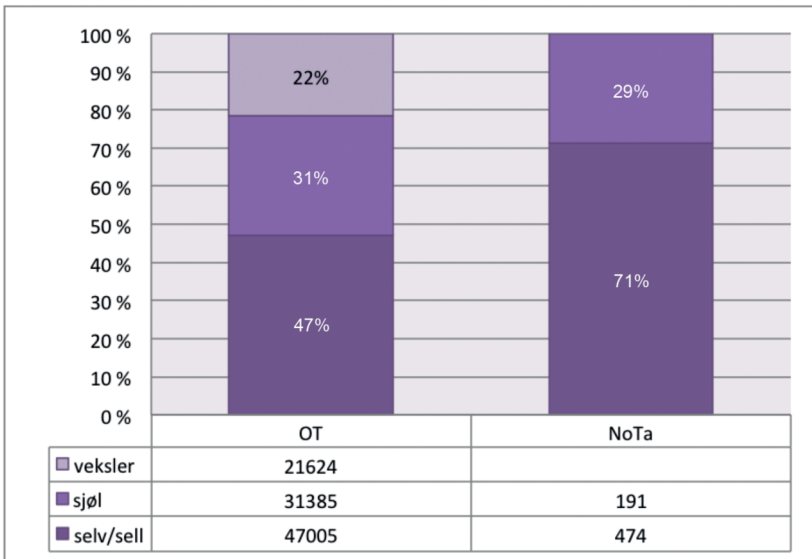


Figur 4: Fordelinga av (h)vordan og åssen for respondentene fra vest i OT²⁷ og informantene vest i NoTa.

Litt færre fra vest i OT rapporterer at de bruker bare *åssen* (11 prosent) enn det som er tilfellet i NoTa (18 prosent). Andelen (*h*)*vordan* er derimot lik, med 64 prosent i begge. Differansen for *veksler* er selvsagt lik som for *åssen*, med 26 prosent i OT, mot 19 prosent i NoTa. Antallet treff på *hvordan* og *åssen* for denne kategorien (vest) i NoTa er relativt lav med bare 80 treff til sammen, så et lavt antall av treff vil gi stort utslag i prosent, det vil si at bare ett treff vil slå mye ut på prosenttallet. Særlig derfor er denne likheten mellom de to undersøkelsene overraskende stor. Det store antallet respondenter som står bak dette prosenttallet i OT, underbygger også tallene fra NoTa. For det er ikke bare for denne variabelen at tallene i NoTa og OT ligger tett. Tendensen er også veldig tydelig for neste variabel vi skal se på. Før vi går videre, vil vi likevel påpeke at *vordan* ble valgt av mange på tross av det avvikende skriftbildet. Det er i likhet med det vi nevnte for *sell* i 5.2.1, evidens for at transkripsjonen vi valgte, antakelig ikke har vært så problematisk. Disse resultatene peker i retning av at svært mange av respondentene har gjort en vurdering av talespråklige former. Det kan i alle fall tyde på at de ikke har gjort en vurdering av denne spesielle transkripsjonen. Som nevnt var dette også en sentral problemstilling knytta til variabelen *sel(l)v* og *sjøl*, som blir presentert i neste avsnitt.

7.3 En grammatisk nøtt: *sel(l)v* og *sjøl*

Resultatene for *sel(l)v/sjøl* var noe uventa. Vi hadde regna med at fles-teparten ville rapportere at de brukte *sell*, og ikke *sjøl*, fordi *selv* er mer frekvent i NoTa (jf. tallene i Figur 6 under). *Selv* er dessuten betydelig mer frekvent i skrift enn *sjøl*, jf. Leksikografisk bokmålskorpus (LBK), der det er 110 839 treff på *selv*, og 3254 treff på *sjøl*, noe som kunne tenkes å ha betydning for statusen til denne variabelen. I tillegg hadde vi en klar oppfatning om at *sell* generelt har høyere prestisje enn *sjøl* på bakgrunn av tilknytninga til talemålet på vestkanten. Likevel var vi forberedt på at det ukjente skriftbildet til *sell*, kunne føre til at testresultatene var lite valide for det formålet vi ønska. Hvis *sjøl* hadde vært den mest populære varianten i OT, på tross av forventet lavere prestisje og at *selv* er mest frekvent i både NoTa og Leksikografisk bokmålskorpus, ville det vært en klar indikasjon på at vi hadde hatt enda større metodologiske utfordringer med testen. Men *sell* var den varianten som flest rapporterte å bruke i OT. I utgangspunktet så fordelinga (mellom både vest og rest) i NoTa og OT slik ut:



Figur 5: Fordelinga av *sjøl* og *sel(l/v)* i OT og NoTa.

Riktignok har vi ikke regna ut hvor mange som veksler i NoTa i denne framstillinga, men fordi det er så stor overvekt av treff på *selv*, er det likevel en umulighet at tallene kan bli like forholdstallene vi ser her fra OT. Dette preliminare resultatet tok vi først til inntekt for at denne variabelen kunne være selve beviset for at dataene måtte vurderes på en annen måte enn i direkte sammenheng med faktisk talespråk, men heller kanskje måtte tolkes mer i retning av en ren holdnings- eller persepsjonsstudie. Men etter å ha hørt gjennom alle treffene på *selv/sjøl* i NoTa, fikk vi en ny oppfatning av dette. *Selv* og *sjøl* brukes nemlig på tre ulike måter i NoTa: som fokuserende adverb i nomen- og adverbfraser (Faarlund, Lie & Vannebo 1997: 337), innledende, alene og sammen med *om*, i ”vedgåingssetninger” (sst.: 1042), og i anaforiske uttrykk. Ut fra treffene i NoTa er det bare når *selv/sjøl* er en anafor, at det er variasjon mellom *selv* og *sjøl* i oslomålet, noe vi kan se i disse eksemplene:

Fokuserende adverb:

NoTa 117: husker jeg foreldrene mine syns var (.) ille for det var så mye trafikk (.) på Sinsenkrysset (.) **selv** den gangen

NoTa 052: *selv* nå etter så mange år

NoTa 074: man hører litt forskjell på dialektene (.) i Sogn og Fjordane
øg (.) *selv* for oss som sitter her borte

“Vedgåingssetninger”:

NoTa 002: det er sier jo sitt *selv om* de faktisk bruker dollar i El
Salvador nå

NoTa 006: *selv med* (.) likt sikkerhetsnivå så er det mye større sjanse
for at e det skjer der enn på Svalbard

NoTa 152: *selv om* fler av kompisene mine gjorde det så blei jeg blei
liksom aldri med på det

Anaforiske uttrykk:

NoTa 34: ja jeg liker det godt *sjøl* jeg

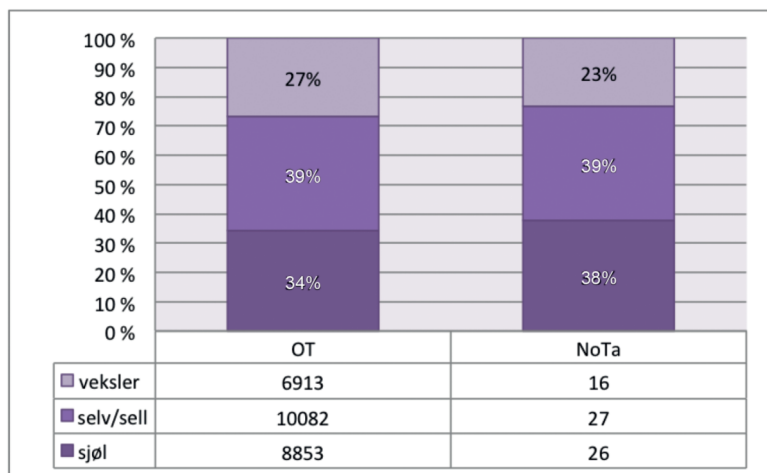
NoTa 103: jeg kan jo ikke huske det *selv* (Og her er det variasjon.)

Vi må påpeke her at *sjøl* blir brukt i en vedgåingssetning to ganger i NoTa. Det er samme informant som bruker det begge gangene, nummer 169. Han er 49 år, oppvokst i Kongsvinger, og har bodd lengst i Ullensaker, og det er nærliggende å tenke seg at 169 bruker en del dialekttrekk som ikke nødvendigvis hører hjemme i Oslo.

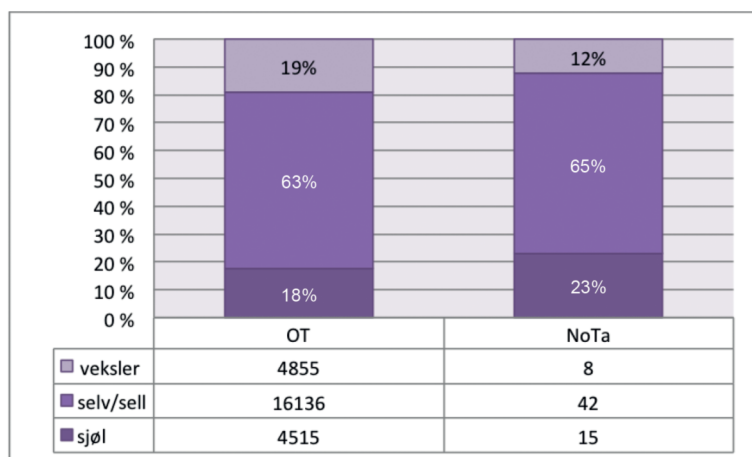
I eksempelsetninga vi brukte i OT, var *sjøl* og *sell* anaforer, nemlig: "Han gjør det SELL"/"Han gjør det SJØL". Da vi plukka ut bare de treffene på *sjøl* og *selv* i NoTa som var anaforer, fikk vi resultatet i Figur 6 og 7 under; se på figur 6 først.

Tallene for NoTa og OT er, i likhet med for *åssen* og *(h)vordan*, like også her. 23 prosent i NoTa veksler mellom *selv* og *sjøl*, mens 27 prosent rapporterer at de gjør det. Når det gjelder varianten *sel(l)v*, bruker 39 prosent av informantene i NoTa bare denne varianten, like stor andel som rapporterer at de gjør det. For *sjøl* er det 38 prosent i rest som bare bruker denne varianten, mens 34 prosent rapporterer at de bare bruker den.

Også i vest ligger tallene fra NoTa og OT nært opptil hverandre, som vist i Figur 7 under.



Figur 6: Fordelinga av *sel(l/v)* og *sjøl* for respondentene fra øst i OT²⁸ og informantene rest i NoTa.



Figur 7: Fordelinga av *sel(l/v)* og *sjøl* for respondentene fra vest i OT²⁹ og informantene vest i NoTa.

I vest mener 19 prosent i OT at de veksler, mens 12 prosent faktisk gjør det i NoTa. Igjen er antall mulige treff i NoTa såpass lavt at små forskjeller vil gjøre store utslag i prosent, men likevel er mønsteret tydelig. I OT mener 63 prosent at de sier *sell*, og 65 prosent i NoTa bruker bare *selv*. 18 prosent i OT mener de bare bruker *sjøl*, mens 23 prosent gjør det i NoTa.

Når vi ser på resultatene fra OT, er det særlig overraskende at den rapporterte bruken av *sell/sjøl* ser ut til å reflektere kun den grammatiske bruken av dette determinativet som framgår i eksempelsetninga i OT. Som vi viser her, kan *sel(l)v* brukes på flere måter i talemålet i Oslo, men *sjøl* ser bare ut til å kunne brukes som en anafor. Resultatene fra OT ser ut til å kun reflektere den bruken som brukes i eksempelsetninga – nemlig den anaforiske bruken. Dette er noe man skulle regne med at krevde en stor grad av grammatisk bevissthet hos respondentene, en bevissthet det er lite sannsynlig at rundt 50 000 beboere i Oslo har. At resultatet likevel ser ut til å reflektere denne bevisstheten, mener vi henger sammen med de store talls lov, eller for å sitere Aristoteles: "The whole is more than the sum of the parts."

8 Oppsummering og avslutning

Med OT har vi vist at det er mulig å få et stort antall mennesker i tale samtidig og på kort tid, samtidig som nye, tekniske løsninger også åpner for at dataene vi samler inn, kan behandles effektivt og på en helt annen måte enn tidligere. Vi har altså vist at Internett og nye tekniske verktøy gir nye muligheter for velprøvde undersøkelsesmetoder som den tradisjonelle spørrelisteundersøkelsen. Det er ikke mange år siden en undersøkelse som OT ville vært nærmest umulig å gjennomføre fordi det ville vært vanskelig å nå så mange respondenter, dessuten ville den enorme datamengden blitt altfor ressurskrevende å håndtere. Den enorme kvantiteten som blir tilgjengeliggjort i OT, er sånn sett relativt ny i språkvitenskapelig sammenheng.

Som vist i analysene av hvordan variablene *assen/(h)vordan* og *sel(l)v/sjøl* fordeler seg i henholdsvis OT og NoTa, er det tydelig at de rapporterte talemålsdataene fra OT i stor grad samsvarer med faktisk språkbruk i NoTa-materialet. Dette er overraskende, ikke minst med tanke på den skepsisen som eksisterer mot å bruke selvrapportering som metode for å få kunnskap om faktisk språkbruk (se for eksempel Labov 1991[1972], Bloomfield 1944). Denne skepsisen er på mange måter forståelig, og årsaken til at dataene fra OT ser ut til å være på linje med dataene fra talespråkkorpuset, henger trolig sammen med det høye antallet respondenter.

Selv om noen osloborgere kanskje har et bevisst forhold til egen talevariasjon, er det det at summen av svar er så stor, som sørger for at det gir et godt oversiktsbilde av språksituasjonen i Oslo. Samtidig må vi også nevne at disse resultatene også styrker troverdigheten til representativiteten til NoTa-innsamlingen, og dermed også utsagnene til Sankoff (1980: 51) om at omtrent 150 informanter er tilstrekkelig i kvantitative undersøkelser for å få oversikt over et språksamfunn. Antallet informanter i NoTa ligger tett opptil det antallet, med 166 informanter.

Dette synliggjør også at ved en masseinnsamling som OT, som er en svært lite ressurskrevende datainnsamling, er utfordringer knytta til testformatet og -situasjonen mindre. Surowiecki (2004) argumenterer for at tekniske hjelpemidler kan tilgjengeliggjøre gruppers kollektive kunnskap eller væremåte, noe resultatene fra OT underbygger. Men hver for seg er ikke nødvendigvis kvaliteten særlig høy på svarene fra OT, i tillegg til at det er ganske få variabler vi har fått undersøkt. Dette står i motsetning til NoTa, der bidragene fra hver og én informant er av høy kvalitet, i tillegg til at NoTa-korpuset kan tilby tilgang til svært mange ulike variabler – noe som altså fordrer en ressurskrevende datainnsamling.

Noter

1. www.tekstlab.uio.no/nota/scandiasyn/index.html (24. april 2014).
2. www.tekstlab.uio.no/cambridge_survey (24. april 2014).
3. Det er selvsagt mulig at respondenter i OT har samarbeida, men dette kommer inn under argumentasjonen i 6.3 om at det er vanskelig for enkeltpersoner eller mindre grupper alene å påvirke resultatene når det er så mange besvarelser.
4. Vi vil presisere at det er en forskjell på språkholdninger generelt og oppfatninger om eller evalueringer av språk spesielt. Røstad (2008: 17) argumenterer for at holdningsbegrepet er mer abstrakt og generelt i motsetning til evalueringsaspektet som gir en mer subjektiv dimensjon (merk også at dette er et viktig moment i Surowieckis argumentasjon, nemlig det individuelle aspektet), og sånn som vi bruker oppfatninger her, er det på linje med det Røstad kaller evalueringer.
5. www.tekstlab.uio.no/nota/oslo/index.html (6. november 2012).
6. Se for eksempel <http://www.nrk.no/norge/problemene-oslo-ikke-klarar-a-lose-1.7996583> (23. april 2014).
7. Takk til Atle Brunvoll i *Aftenposten* som har vært hjelpsom og vist stor interesse for prosjektet.
8. Dette gjelder for de første 11 variablene vi testa i OT. De neste 11 var knytta til språkbruk i flerspråklige og flerkulturelle miljøer. Se forøvrig Ims (2013 og u.a.).
9. Denne avgjørelsen kan illustrere hvor fort teknologien utvikler seg. Allerede nå, fire år etterpå, virker ikke denne problemstillingen like relevant lenger.
10. www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/index.html (24. september 2012).

11. Det kan være grunn til å diskutere tekstsjangrene i LBK i denne sammenhengen, men det vil vi ikke gå nærmere inn på her.
12. www.ssb.no/vis/emner/07/02/30/medie/art-2012-03-20-01.html (18. mai 2010).
13. www.ssb.no/vis/magasinet/slik_lever_vi/art-2012-03-20-01.html (18. mai 2010).
14. Disse tallene er kun basert på de respondentene i OT som oppgir at de bor i Oslo.
15. Dette tallet er beregna på Oslos befolkning mellom 13 og 19 år. Barn som er 12 år og yngre er ikke tatt med i denne beregninga.
16. Siden Oslo-statistikken deler opp kategoriene i 50–66 år, og 67–79 år, har vi gjort en beregning for kategoriene 50–59 år og 60 < år.
17. www.utviklings-og kompetansestaten.oslo.kommune.no/getfile.php/utviklings%20og%20kompetansestaten%20%28UKE%29/Internett%20%28UKE%29/Dokumenter/Oslostatistikken/Befolkning/Folkemengde/Folkemengde%202012/02.04befdelbydel2012.htm (18. mai 2012).
18. Mye av potensialet for metodeutviklinga ligger nettopp her, nemlig hva denne typen undersøkelser egner seg til og ikke.
19. Om man kan sies å teste noe utover akkurat den variabelen man faktisk bruker i en test som OT, er en egen diskusjon vi ikke tar i denne sammenhengen.
20. Det kan diskuteres om *hvorledes* også kan brukes på samme måte som *hvordan* og *åssen*, men mens *hvorledes* kun hadde to treff i TAUS, fins det ingen treff på *hvorledes* i NoTa, så vi regner ikke *hvorledes* som en levedyktig del av oslomålet lenger.
21. Også for variabelen *vet/veit* kan det hende at vi vil finne liknende resultater som for *hvordan/åssen* og *selv/sjøl*, men det har vi ikke sett nærmere på her.
22. www.tekstlab.uio.no/nota/oslo/ (10. mai 2012).
23. *Øst* vil si bydelene Alna, Bjerke, Gamle Oslo, Grorud, Grünerløkka, Østensjø, Søndre Nordstrand, Sagene, Sentrum og Stovner, i tillegg til respondentene som bare oppgir "Oslo øst" (25 848 stykker). *Vest* vil si bydelene Frogner, Nordre Aker, Nordstrand, St. Hans-haugen, Ullern og Vestre Aker, i tillegg til respondentene som bare oppgir "Oslo vest" (25 506 stykker).
24. www.tekstlab.uio.no/nota/oslo/ (23. april 2012).
25. Svendsen (2012) problematiserer også dette.
26. Tallene fra materialet der vi kun har tatt med første gangs registrering av IP-adressene, er: *åssen*: 21 prosent, *(h)vordan*: 44 prosent, *veksler*: 35 prosent.
27. Tallene fra materialet der vi kun har tatt med første gangs registrering av IP-adressene, er: *åssen*: 11 prosent, *(h)vordan*: 63 prosent, *veksler*: 26 prosent.
28. Tallene fra materialet der vi kun har tatt med første gangs registrering av IP-adressene, er: *sjøl*: 34 prosent, *sell*: 39 prosent, *veksler*: 27 prosent.
29. Tallene fra materialet der vi kun har tatt med første gangs registrering av IP-adressene, er: *sjøl*: 18 prosent, *sell*: 63 prosent, *veksler*: 19 prosent.

Referanser

- Aalen, Odd O. 1998: *Innføring i statistikk med medisinske eksempler*. Oslo: Ad Notam Gyldendal.
- Akselberg, Gunnstein 2008: "Kor pent bergensk snakkar du?" ei gransking av sjølvrapporterte språkdata til avisa Bergensk Tidende si heimeside,

- bt.no. *Målbryting. Skrifter frå Norsk sociolingvistisk nettverk* (SONE) 9, 149–166.
- Alnæs, Karsten 1963: *Hovedlinjer av Oslo-dialektens utvikling i det siste hundreåret*. Hovedoppgave. Universitetet i Oslo.
- Aure, Marit, Bjørg Langset & Kjetil Sørli 2011: Flyttemotiver og bostedsvalg. Nye muligheter til å forstå norske by- og flytteprosesser. *Plan. Tidskrift for samfunnsplanlegging, bolig og byplan og regional utvikling* 5, 42–48.
- Benum, Edgeir 2000 [1994]: *Byråkratienes by*. Bind 5. *Oslo bys historie*. Oslo: Cappelen.
- Best, Samuel & Brian Krueger 2004: Internet Data Collection. *Quantitative Applications in the Social Sciences* 141. Sage Publications.
- Bloomfield, Leonard 1944: Secondary and Tertiary Responses to Language. *Language* 2. 20, 45–55.
- Bråthen, Magne mfl. 2007: *Levekår på vandring. Velstand og marginalisering i Oslo*. Fafo-rapport 05. [Oslo]: Fafo.
- Chambers, J.K. & Peter Trudgill 1998 [1980]: *Dialectology*. Second Edition. Cambridge: Cambridge University Press.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo 1997: *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Hanssen, Eskil mfl. 1978: *Oslomål. Prosjektbeskrivelse og syntaktisk analyse av oslomål med henblikk på sosiale skilnader*. Skrift nr. 6. Talemålsundersøkelsen i Oslo. Oslo: Novus.
- Hanssen, Eskil 1986: Ordforrådet i et talemålsmateriale. Hanssen, Eskil mfl. (red.): *Artikler 1–4 Talemålsundersøkelsen i Oslo* (TAUS). Skrift nr. 1–4. Talemålsundersøkelsen i Oslo. Oslo: Novus, 29–67.
- Hanssen, Eskil 2010: *Dialekter i Norge*. Bergen: Fagbokforlaget.
- Hårstad, Stian 2008: Store tal frå trondheimske tale. Metodologisk drøfting av ei nettbasert dialektgransking. *Målbryting. Skrifter frå Norsk sociolingvistisk nettverk* (SONE) 9, 127–147.
- Hårstad, Stian 2010: *Unge språkbrukere i gammel by. En sociolingvistisk studie av ungdoms talemål i Trondheim*. Ph.d.-avhandling. Trondheim: INL, NTNU.
- Ims, Ingunn Indrebø 2013: Språklig registerdanning og verditilskrivning. Betegnelser på nye måtar å snakke norsk på i Oslo. *NOA - norsk som andrespråk* 2, 37–71.
- Ims, Ingunn Indrebø u.a.: Doktoravhandling for graden ph.d. ILN, Universitetet i Oslo.

- Jahnsen, Vanja 2001: *Øst og vest for elva. En sosiolingvistisk undersøkelse av talemålet i Oslo*. Hovedoppgave. Oslo: Universitetet i Oslo.
- Jahr, Ernst Håkon 1986: L-fonema i Oslo bymål. Hanssen, Eskil mfl. (red.): *Artikler 1-4. Talemålsundersøkelsen i Oslo (TAUS)*. Oslo: Novus, kap. 1, 9–27.
- Johannessen, Janne Bondi 2008: Oslospråket i tall. Johannessen, Janne Bondi & Kristin Hagen (red.): *Språk i Oslo. Ny forskning omkring talespråk*. Oslo: Novus, 235–242.
- Johnson, Keith 2008: *Quantitative methods in Linguistics*. Malden: Blackwell.
- Kjeldstadli, Knut 2000 [1994]: *Den delte byen*. Bind 4. *Oslo bys historie*. Oslo: Cappelen.
- Kristiansen, Tore & Nikolas Coupland 2011: SLICE: Critical perspectives on language (de)standardisation. Kristiansen, Tore & Nikolas Coupland (red.): *Standard Languages and Language Standards in a Changing Europe*. [Oslo]: Novus, 11–35.
- Kristoffersen, Gjert & Hanne Gram Simonsen 2008: Osjlo! En undersøkelse av uttalen av sl-sekvenser i NoTa-korpuset. Johannessen, Janne Bondi & Kristin Hagen (red.): *Språk i Oslo. Ny forskning omkring talespråk*. Oslo: Novus, 96–108.
- Labov, William 1991[1972]: *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Larsen, Amund B. 1907: *Kristiania Bymål. Vulgærsproget med henblikk på den utvungne dagligtale*. Kristiania: Bymålslaget.
- Lødrup, Helge 2011: Hvor mange genus er det i Oslo-dialekten? *Maal og minne* 2, 120–136.
- Mordal, Tove L. 1989: *Som man spør får man svar. Arbeid med survey-opplegg*. Spydeberg: Tano.
- Niedzielski, Nancy A. & Dennis R. Preston 2000: *Folk Linguistics*. New York: Mouton de Gruyter.
- Opsahl, Toril & Unn Røyneland 2009: Osloundom – født på solsiden eller i skyggen av standardtalemålet? *Norsk Lingvistisk Tidsskrift* 1, 95–119.
- Preston, Dennis R. 1989: *Perceptual Dialectology. Nonlinguists' Views of Areal Linguistics*. Dordrecht: Foris.
- Røstad, Rune 2008: Språkoppfatningar og sosiolingvistikken. *Målbryting. Skrifter frå Norsk sosiolingvistisk nettverk (SONE)* 9, 7–33.
- Sankoff, Gillian 1980: *The Social Life of Language*. Philadelphia: Pennsylvania Press.

- Stjernholm, Karine 2013: *Stedet velger ikke lenger deg – du velger et sted. Tre artikler om språk i Oslo*. Ph.d.-avhandling. [Oslo]: ILN, Universitetet i Oslo.
- Surowiecki, James 2004: *The Wisdom of Crowds*. New York: Doubleday.
- Svendsen, Bente Ailin 2012: Et språklig bakholdsangrep? “Østfold-l-ens” inntog i Oslo. Enger, Hans Olav & Unn Røyneland (red.): *Fra holtijaR til holting. Språkhistoriske og språksosiologiske artikler til Arne Torp på 70-årsdagen*. Oslo: Novus, 349–365.

Korpus

NoTa: Norsk talespråkskorpus – Oslodelen. Tekstlaboratoriet, ILN, Universitetet i Oslo. <http://www.tekstlab.uio.no/nota/oslo/index.html>

English Summary

The article discusses methodological and theoretical challenges of using the Internet to gather language data. During two months in 2010, the electronic survey *The Oslo test* (OT) gathered over 115,000 answers about variation in the Oslo dialect. The OT had similarities with a classical questions list survey, but using the Internet as medium for data collection is relatively new in language research. Since the method has not been used extensively before, there is no established theoretical framework that can be used as foundation for the analysis of these data. The test situation was particularly inaccessible to the researcher and the OT generated far more answers than established theoretical frameworks account for. The article presents therefore mainly (1) a suggestion for a theoretical framework for the analysis, and (2) a comparison of the self-reported data with actual speech recordings from Norwegian Speech Corpus - the Oslo part (NoTa).

Karine Stjernholm
Von der Lippes gate 6
0454 Oslo
karine.stjernholm@gmail.com

Ingunn Indrebø Ims
Solfjellveien 32
1389 Heggedal
ingunn.ims@gmail.com